Catalogue no. 82-003-X
ISSN 1209-1367

Health Reports

Two approaches to linking census and hospital data

by Michelle Rotermann, Claudia Sanmartin, Gisèle Carrière, Richard Trudeau, Hélène St-Jean, Abdelnasser Saïdi, Alexander Reicker, Aimé Ntwari and Eric Hortop

Release date: October 15, 2014

Statistics Canada Statistique Canada

Canada

How to obtain more information

For information about this product or the wide range of services and data available from Statistics Canada, visit our website, www.statcan.gc.ca.

You can also contact us by

email at infostats@statcan.gc.ca,

telephone, from Monday to Friday, 8:30 a.m. to 4:30 p.m., at the following toll-free numbers:

- Statistical Information Service 1-800-263-1136
- National telecommunications device for the hearing impaired 1-800-363-7629
- Fax line 1-877-287-4369

Depository Services Program

- Inquiries line 1-800-635-7943
- Fax line 1-800-565-7757

To access this product

This product, Catalogue no. 82-003-X, is available free in electronic format. To obtain a single issue, visit our website, www.statcan.gc.ca, and browse by "Key resource" > "Publications."

Standards of service to the public

Statistics Canada is committed to serving its clients in a prompt, reliable and courteous manner. To this end, Statistics Canada has developed standards of service that its employees observe. To obtain a copy of these service standards, please contact Statistics Canada toll-free at 1-800-263-1136. The service standards are also published on www.statcan.gc.ca under "About us" > "The agency" > "Providing services to Canadians."

Published by authority of the Minister responsible for Statistics Canada

© Minister of Industry, 2014

All rights reserved. Use of this publication is governed by the Statistics Canada Open Licence Agreement (http://www.statcan.gc.ca/reference/licence-eng.htm).

Cette publication est aussi disponible en français.

Note of appreciation

Canada owes the success of its statistical system to a long-standing partnership between Statistics Canada, the citizens of Canada, its businesses, governments and other institutions. Accurate and timely statistical information could not be produced without their continued co-operation and goodwill.

Standard symbols

The following symbols are used in Statistics Canada publications:

.	not available for any reference period
..	not available for a specific reference period
...	not applicable
0	true zero or a value rounded to zero
0^s	value rounded to 0 (zero) where there is a meaningful distinction between true zero and the value that was rounded
^{p}	preliminary
^{r}	revised
x	suppressed to meet the confidentiality requirements of the *Statistics Act*
^{E}	use with caution
F	too unreliable to be published
*	significantly different from reference category ($p < 0.05$)

Two approaches to linking census and hospital data

by Michelle Rotermann, Claudia Sanmartin, Gisèle Carrière, Richard Trudeau, Hélène St-Jean, Abdelnasser Saïdi, Alexander Reicker, Aimé Ntwari and Eric Hortop

Abstract

Background

This study compares registry and non-registry approaches to linking 2006 Census of Population data for Manitoba and Ontario to hospital data from the Discharge Abstract Database (DAD).

Data and methods

Using a probabilistic linkage, the *registry* approach linked the census data to provincial health insurance registries, followed by a deterministic linkage to the DAD based on health insurance number (HIN). The *non-registry* approach used hierarchical deterministic exact matching based on three variables common to both files to link census data to the DAD. The approaches were compared in terms of linkage and coverage rates, sensitivity and specificity, and consistency of HINs on the linked records.

Results

Results of the registry and non-registry linkage approaches were similar. In Manitoba, 7% and 6% of census long-form respondents linked to the DAD with the registry and non-registry linkage approaches, respectively; in Ontario, the linkage rate was 5% for both approaches. With the registry approach, the linked census-DAD data represented 84% (weighted) of hospital admissions in the 2006/2007 DAD in both provinces, compared with 82% in Manitoba and in Ontario with the non-registry approach.

Interpretation

In the absence of access to provincial health insurance registries with which census data can be linked, a non-registry approach can be used to create a research-quality dataset.

Keywords

Coverage, data collection, data linkage, databases, medical records, hospital records, meta-analysis

Authors

Michelle Rotermann (michelle.rotermann@statcan.gc.ca), Claudia Sanmartin and Gisèle Carrière are with the Health Analysis Division, Richard Trudeau is with the Health Statistics Division, and Hélène St-Jean, Abdelnasser Saïdi, Alexander Reicker, Aimé Ntwari and Eric Hortop are with the Household Survey Methods Division at Statistics Canada, Ottawa, Ontario K1A 0T6.

Record linkage, the process of matching records across or within datasets, is common in health research.[1-7] The goal is to create an enriched dataset with wider applications.[6-11] The data suited for linkage are those that are complementary—information unavailable in one source is available in the other.

Accurate data linkage depends on the availability of co-occurring unique identifiers.[12,13] Each identifier should pertain to only one person, and each person should have only one identifier.[14] For health-related linkages in Canada, provincial health insurance numbers (HINs) have been used.[2,3,11,15] However, in most databases (for instance, vital statistics, census and tax files), HINs are not available.

One option is a *registry* approach to linking databases. Provincial health insurance registries can act as "bridge" files for linkage because they contain HINs along with names and other identifying variables.[2,3,11,15] However, registries are not always available. Alternatively, a *non-registry* approach can be taken. This involves matching records in different databases, using combinations of co-occurring person-level information, such as birth date and postal code.[13,14,16,17]

If databases pertain to the same population, it is generally expected that most records will link. But if only a fraction of records is expected to link, determining a reasonable linkage rate *a priori* is problematic. This is typical of health-based linkages—for example, limited numbers of individuals will be hospitalized or die during follow-up. In such situations, linkages have been evaluated by comparing the results of different approaches using the same data[12,14,18-20]; by comparing outcome rates and percentage distributions of variables available in linked and unlinked data[14,21,22]; and by calculating sensitivity and specificity.[19,23,24] Based on Canadian studies, match rates around 75% among records expected to link are considered appropriate for research applications.[5,25,26]

This study compares a *registry* and *non-registry* approach to linking the 2006 Census of Population with hospital data from the Discharge Abstract Database (DAD) for Manitoba and Ontario, two provinces for which health insurance registry data (HINs) are available to Statistics Canada. The aim is to determine if a research-quality dataset can be produced without the aid of "bridge" data from provincial health insurance registries. The linkage was approved by Statistics Canada's Policy Committee,[27] and was governed by the Record Linkage Directive.[28]

Data and methods

Data sources

2006 Census of Population

The 2006 Census collected information using both short- and long-form questionnaires. The entire population answered the 7 basic demographic questions on the short form, which included the birth date, sex and postal code of all members of each household.[29] As well, about 20% of private households were randomly selected to answer the long-form questionnaire, which asked 52 additional questions about income, education, ethnicity, Aboriginal status, etc.[29]

The census short-form data file includes long-form respondents, and so contains records pertaining to nearly the entire population (97% in Manitoba; 96% in Ontario). The long-form data file pertains only to the 20% of households selected to provide in-depth information.

Before linkage, the census file was cleaned, unduplicated and validated.[29,30] The short-form file was linked to the provincial health insurance registries and then to the DAD (*registry* approach), or directly to the DAD (*non-registry* approach). Inclusion of all census records in the linkage made it possible to identify and remove provincial health insurance registry and DAD records that pertained to short-form-only respondents. Only records pertaining to people who completed a long-form census questionnaire were included in the study cohort.

Provincial health insurance registries (registry approach)

For the *registry* approach, the Manitoba Health Services Insurance Plan (MHSIP) Registration File and the Ontario Registered Persons Database (RPDB) were used as "bridge" files. Linking the census data to a provincial health insurance registry makes it possible to then link the census data to the DAD, based on HIN agreement.

The MHSIP and RPDB contain records for all individuals registered to receive health services in Manitoba and Ontario, respectively. Because registrants are not obliged to pay, population coverage is high.[31-33] Records for people no longer living in the province, but who retain their coverage for up to three months after moving, are included in both registries. New residents must wait three months for coverage. People covered by other plans (for instance, prison inmates, members of the RCMP and Canadian Forces) are excluded.

Before linkage to the census data, the MHSIP and RPDB were pre-processed, including identification of people with multiple HINs (Manitoba = 0.2% or 3,588; Ontario = 1% or 165,123). Only records with surnames, birth dates before January 1, 2007, insurance coverage effective between December 31, 2005 and January 1, 2007, and where applicable, death dates after December 31, 2005, were eligible for the study cohort.[36]

2006/2007 Discharge Abstract Database

The DAD contains demographic, administrative (including HIN) and clinical data for all acute-care and some psychiatric, chronic rehabilitation and day-surgery hospital discharges.[34-35] The 2006/2007 version pertains to hospital discharges from April 1, 2006 through March 31, 2007 (n = 3,186,079).

2005/2006/2007 T1 Personal Master Files (non-registry approach)

To account for changes in postal codes over time and to improve the chances of linkage to the DAD using the *non-registry* approach, postal codes from Statistics Canada's 2005, 2006 and 2007 tax files (T1 Personal Master Files - T1PMF) were added to the short-form census file. Based on sex, birth date, and partial family and given names, most census records (91%) linked to at least one year of tax data. For people who did not file taxes annually and/or were not required to (for instance, children), postal codes were identified and assigned using information from other household members who were tax-filers.

Record linkage

Registry approach

The *registry*-based linkage was conducted in two steps: 1) Manitoba and Ontario short-form census records were probabilistically linked to the provincial health insurance registries to obtain HINs; and 2) based on HINs, the registry-linked long-form census records were linked deterministically to the DAD (Figure 1).

Probability scores based on similarities of birth date, postal code, sex, surname, and given name were used to estimate the likelihood that linking records represented the same person.[19,37,38] Weights (positive/negative) were assigned to the comparison fields, which were summed to create a total linkage weight. Separate thresholds to distinguish true matches from non-matches were pre-determined for Manitoba and Ontario based on distributions of these weights. Because of the size of the Ontario files, comparing every census record with every registry record was prohibitive. Consequently, the Ontario file was split, so that only records sharing the same sex were compared. This was not done for Manitoba, because the population and files were smaller.

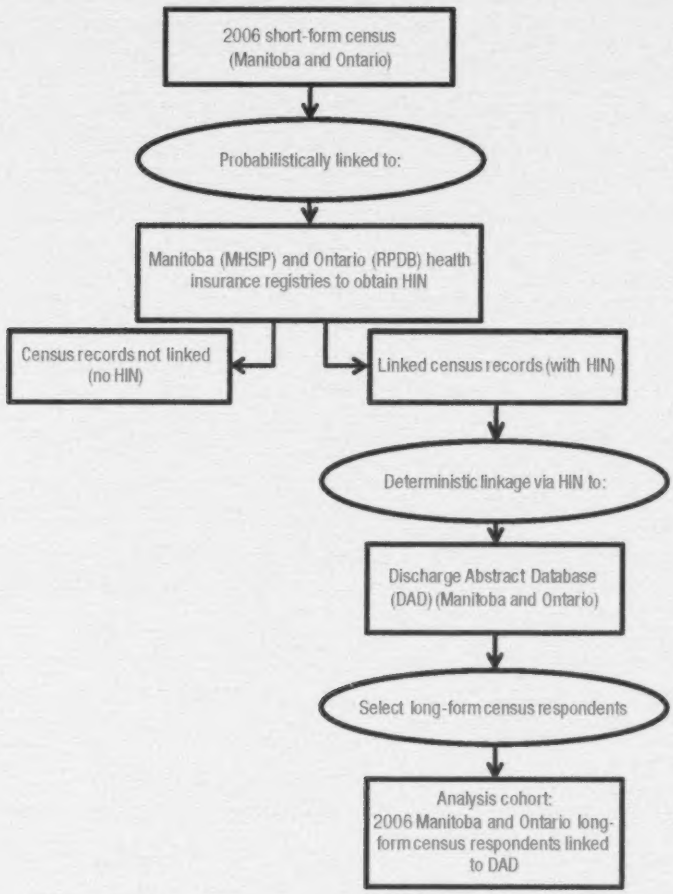
Record-pairs were ranked. Those scoring above the pre-determined threshold were accepted as matches. Before the thresholds were finalized, record-pairs ranking close to them were reviewed, and thresholds were adjusted where necessary.

The linked census-registry files containing both the census identifiers and HIN were then used to deterministically link the census identifiers to the DAD files. Only census identifiers corresponding to long-form census respondents were retained for analysis.[38]

Non-registry approach

The *non-registry* approach used hierarchical deterministic exact matching on linkage keys comprised of combinations of three variables (birth date, postal code and sex) common to the census

Figure 1
Registry-based linkage, Manitoba and Ontario


2006 short-form census (Manitoba and Ontario)
Probabilistically linked to:
Manitoba (MHSIP) and Ontario (RPDB) health insurance registries to obtain HIN
Census records not linked (no HIN)
Linked census records (with HIN)
Deterministic linkage via HIN to:
Discharge Abstract Database (DAD) (Manitoba and Ontario)
Select long-form census respondents
Analysis cohort: 2006 Manitoba and Ontario long-form census respondents linked to DAD

short-form and DAD records (Figure 2). Matching involved comparing census-DAD pairs to determine if they referred to the same person. If only matches using one key had been accepted, linkage rates would be reduced.[12] By using multiple keys in succession, hierarchical deterministic exact matching is a refinement that maximizes the discriminating power of the linking information and minimizes the impact of missing values and errors.[16]

Data were reformatted in preparation for linkage. When records did not indicate sex (n ~ 312,000 census; n ~ 300 DAD), the existing record was assigned a sex, and an additional record was created with the same birth date and postal code, but the opposite sex. When the original census postal code and post-processed postal code differed (n~525,000), an additional record was also created.

Multiple records pertaining to the same person were identified in the census and DAD files based on GroupID and HINKey (HIN+Province). This facilitated their removal after the linkage was completed.

The *non-registry* linkage used an iterative approach that applied 28 rules to the census-DAD files in succession. Early iterations observed stringent rules; later passes tolerated divergence. For example, the first iteration required an exact match in the census and DAD records on birth date, sex and postal code. Iterations 2 to 4 required an exact match on birth date, sex and postal code in the DAD and T1PMF files for 2005, 2006 and 2007. Iterations 5 to 10 relaxed the rules for postal code, allowing one of the six characters to be dropped. This process was repeated using the T1PMF 2005, 2006 and 2007 postal codes (iterations 11 to 28). After each iteration, census records with the same GroupID and DAD records with the same HINKey as those that linked were removed from future iterations to ensure that people who were represented in the census and DAD files because of multiple linkage keys were linked only once.

Records with duplicate linkage keys could exist in each dataset, particularly the DAD, given that in a single year one person could have multiple hospital records with the same birth date, sex and postal code. To improve efficiency and remove the possibility of ties, the census and DAD files were unduplicated using this linkage key prior to linking. After the linkage was completed, the census records that had been added to allow for inconsistent census postal codes and/or missing sex were removed, and the dropped DAD records (multiple hospitalizations of the same person) were added back to the file. Finally, only records of hospitalizations of Manitoba and Ontario residents were retained for this comparison study.

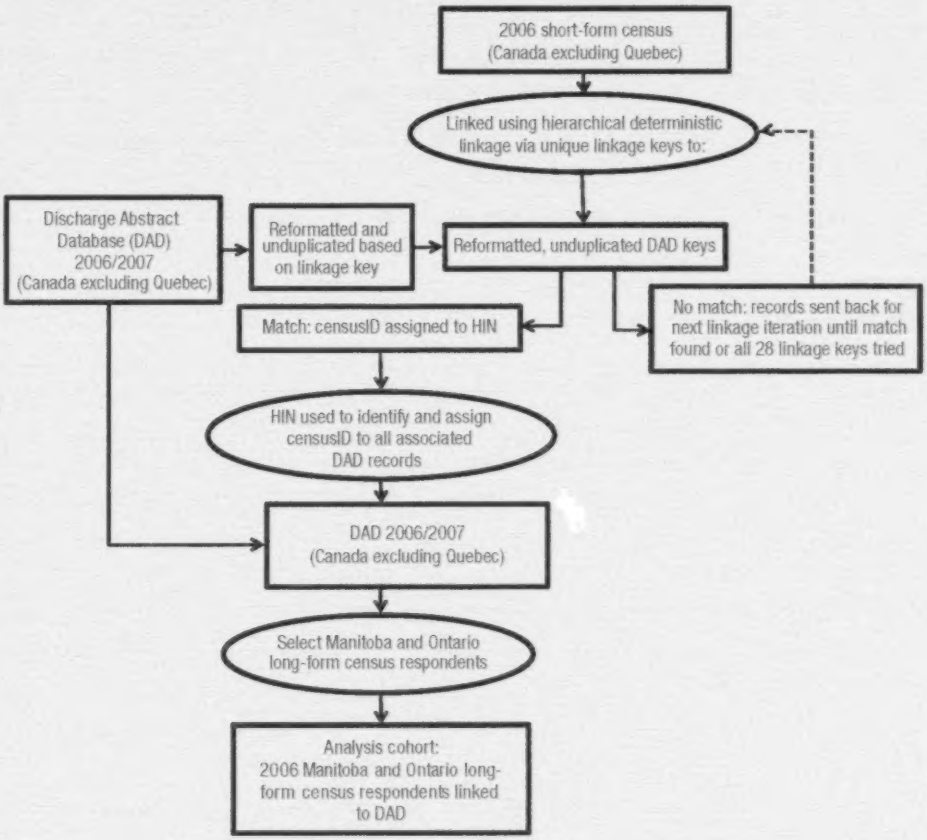
Protecting respondent privacy

Statistics Canada ensures respondent privacy during the linkage process and subsequent use of linked files. Only employees directly involved in the linkage process have access to the unique identifying information required for linkage (such as names and health insurance numbers) and do *not* access health-related information. Once the data linkage process is completed, an analytical file is created from which the identifying information is removed. This de-identified file is accessed by analysts for validation and analysis.

Record counts

For the *registry* approach, 1,111,133 Manitoba and 11,704,729 Ontario short-form census records were sent to linkage with 1,201,152 MHSIP and 13,121,593 RPDB records. Of these, 246,578 of the 278,937 long-form Manitoba and 2,136,455 of the 2,387,911 long-form Ontario census records linked to a health insurance file, and thus, were in-scope for the *registry*-based linkage to the DAD (Table 1).

Figure 2
Non-registry-based linkage, Canada excluding Quebec


2006 short-form census (Canada excluding Quebec)
Linked using hierarchical deterministic linkage via unique linkage keys to:
Discharge Abstract Database (DAD) 2006/2007 (Canada excluding Quebec)
Reformatted and unduplicated based on linkage key
Reformatted, unduplicated DAD keys
No match: records sent back for next linkage iteration until match found or all 28 linkage keys tried
Match: censusID assigned to HIN
HIN used to identify and assign censusID to all associated DAD records
DAD 2006/2007 (Canada excluding Quebec)
Select Manitoba and Ontario long-form census respondents
Analysis cohort: 2006 Manitoba and Ontario long-form census respondents linked to DAD

Because provincial health insurance registries were not required for the *non-registry* approach, most short-form census records for residents of most provinces/territories were eligible for linkage to the DAD (n = 23,592,671). Census records were excluded if the postal code or birth date was invalid or missing, or if both the original census postal code and post-censal postal code were from Quebec. (Quebec census respondents were not eligible for linkage because Statistics Canada does not have access to Quebec inpatient hospital data.) The final *non-registry*-based study cohort consisted of 278,937 Manitoba and 2,387,911 Ontario long-form census records.

Both linkage approaches excluded DAD records with birth dates after Census Day (May 16, 2006). Records for out-of-country residents and for stillbirths were also removed. For the *registry*-based linkage, only DAD records with a valid Manitoba (227,069) or Ontario (1,081,443) HIN were retained. For the *non-registry* linkage, only DAD records with non-missing dates of birth and postal codes were retained (2,106,104).

Validation

Linkage rates

To evaluate the linkage rates achieved with the *registry* approach, the percentages of census respondents linking to either the MHSIP or RPDB were examined. Given differences in the population coverage of the census and provincial health insurance registry files, linkage rates should approach, but not equal, 100%.

For both the *registry* and *non-registry* approaches, the percentages of Manitoba and Ontario long-form census respondents linking to the DAD were calculated overall and by selected socio-demographic characteristics. These linkage rates should reflect the prevalence of being hospitalized at least once in the 2006/2007 fiscal year, and are expected to be higher among groups more likely to be hospitalized (for example, seniors).

Linkage accuracy

Sensitivity (true positives) and specificity (true negatives) were calculated to assess linkage accuracy at the record level. The *registry*-based results were used as the "gold standard" against which the *non-registry* results were compared.[2,3,11,15]

For the census records that linked to the DAD in both approaches, HINs were compared to assess internal consistency. Matches were taken as evidence of correct links.

Coverage analysis

The coverage rates for each approach were calculated by dividing the number of acute-care hospital discharges of long-form census respondents in Manitoba and Ontario using the linked census-DAD data (numerator) by the number of acute-care hospital discharges reported in the 2006/2007 unlinked DAD data (denominator). Unweighted and weighted coverage rates were calculated for each approach overall and for hospitalizations attributable to three "most responsible" diagnoses—circulatory system diseases; injury and poisoning; and pregnancy, childbirth and the puerperium.

Weights are necessary to make inferences at the population level.[39] When weights are applied, coverage rates should approach, but not equal, 100%. When weights are not applied, coverage rates should approach the percentage of the population completing the long-form census (25% in Manitoba; 20% in Ontario). Owing to differences in the populations covered by the linked census-DAD data and the unlinked DAD data, exact agreement is not expected. For instance, the institutionalized population, who are high users of hospital services,[40,41] is represented in the unlinked DAD data, but not in the linked census-DAD data. Census weights were not adjusted for such anomalies, so applying the weights may distort some estimates.

To more closely match the target population of the linked files, DAD records for the following populations were removed from the denominator: residents of seniors' homes (2,368 Manitoba and 24,487 Ontario DAD discharges), people born after Census Day, stillbirths, and non-Canadians.

Socio-demographic characteristics

For each economic family or unattached individual, total after-tax income from all sources and from all family members was summed, adjusted for family composition and size, and separated into *income quintiles*.

Highest level of education for people aged 18 or older was categorized as secondary school graduation, or less than secondary graduation.

Employment status was defined as employed, unemployed or not in the labour force.

Respondents were divided into four categories based on their self-reported knowledge of Canada's *official languages*: English, French, both, or neither.

Information on *Aboriginal status* was derived from the question: "Is this person an Aboriginal person, that is, North American Indian, Métis or Inuit (Eskimo)?" Respondents marked all that applied. Responses were grouped into six categories: North American Indian (only), Métis (only), Inuit (only), other Aboriginal (multiple or indeterminate), Aboriginal (total of four preceding categories), and non-Aboriginal.

Country of birth, citizenship, and immigration status were combined into an *immigrant status* variable: immigrant, non-immigrant or non-permanent resident.

A *one-year residential mobility* variable was created to reflect address changes: same address, moved within Canada, or moved from outside Canada. This was derived by comparing each respondent's municipality and province of residence on Census Day and one year earlier.

A *rural-urban* variable reflected residence location and community size. Farm or non-farm residences in areas with a population of less than 1,000 were considered rural/farm. Other residences were categorized as being in centres with small (1,000 to 29,999), medium (30,000 to 99,999) or large (100,000 or more) populations.

Hospital outcomes

Based on the DAD, the total number of *all-cause hospitalizations* with discharge dates from April 1, 2006 through March 31, 2007 was determined for Manitoba and Ontario.

Hospitalizations attributable to *circulatory system disease* required a most-responsible diagnosis (MRDx) coded to J00-J99 using the *International Classification of Diseases, 10th Revision*. *Injury and poisoning* hospitalizations were those coded S00-S99 or T00-T98. *Pregnancy, childbirth and the puerperium* hospitalizations were those coded O00-O99.

Linkage results

Census-to-provincial health insurance registry (registry approach)

For the *registry* approach, 88% (246,578) of Manitoba long-form census respondents linked to the MHSIP, and 89% (2,136,455) of Ontario long-form census respondents linked to the RPDB, and thus, constituted the study cohort eligible for registry-based linkage to the DAD (Table 1). (The percentages of short-form census respondents linking to the registries were somewhat higher: 93% for Manitoba and 90% for Ontario.)

Long-form census respondents' linkage rates to the health insurance registries differed by socio-demographic characteristics. For example, rates in Manitoba ranged from 85% for children younger than age one to 92% at ages 65 to 74. Linkage rates were relatively low for people in the lowest income quintile, those reporting no knowledge of Canada's official languages, non-permanent residents, and people who had not lived in Canada the year before the census. Linkage rates among Aboriginal groups ranged from 76% to 89%.

Census-to-DAD (non-registry approach)

With the *non-registry* approach, 80% or 1.69 million DAD keys (Canada excluding Quebec) were linked to a short-form census record (Table 2). The majority of links between the census and

Table 1
Number and percentage of 2006 Census respondents linking to Manitoba Health Services Insurance Plan (MHSIP) or Ontario Registered Persons Database (RPDB), by selected socio-demographic characteristics, Manitoba and Ontario, 2006/2007

	Manitoba Census respondents			Ontario Census respondents		
	Total	Linking to MHSIP		Total	Linking to RPDB	
Characteristics	'000	'000	%	'000	'000	%
Total short-form	**1,111.1**	**1,038.9**	**93**	**11,704.7**	**10,578.0**	**90**
Total long-form	**278.9**	**246.6**	**88**	**2,387.9**	**2,136.5**	**89**
Sex						
Male	138.2	120.9	87	1,169.0	1,043.5	89
Female	140.8	125.7	89	1,219.0	1,092.9	90
Age group						
Younger than 1	4.1	3.5	85	26.1	23.9	91
1 to 4	15.9	13.6	86	107.3	97.9	91
5 to 9	21.1	18.3	87	143.4	131.1	91
10 to 14	23.1	20.2	87	163.2	148.2	91
15 to 19	22.2	19.4	87	166.6	147.9	89
20 to 24	19.3	16.2	84	155.4	133.7	86
25 to 34	34.6	30.2	87	299.4	261.3	87
35 to 44	38.8	34.6	89	377.5	335.6	89
45 to 54	39.2	35.4	90	369.4	331.5	90
55 to 64	28.2	25.7	91	269.9	245.3	91
65 to 74	17.1	15.6	92	171.2	155.5	91
75 or older	15.4	13.8	90	138.3	124.5	90
Economic family income quintile						
1 (lowest)	55.8	45.5	82	477.4	401.8	84
2	55.8	48.2	86	477.7	425.5	89
3	55.8	49.9	89	477.6	432.7	91
4	55.8	51.3	92	477.6	436.6	91
5 (highest)	55.8	51.6	93	477.6	439.8	92
Education (aged 18 or older)						
Secondary graduation or more	135.6	122.8	91	1,494.9	1,338.0	90
Less than secondary graduation	65.1	56.0	86	350.9	306.2	87
Employment status						
Employed	127.1	114.7	90	1,216.2	1,089.9	90
Unemployed	10.4	8.9	85	84.9	74.3	87
Not in the labour force	141.4	123.0	87	1,086.8	972.3	89
Knowledge of official languages						
English only	254.2	224.5	88	2,053.7	1,841.3	90
French only	0.4	0.4	89	9.5	8.5	90
Both English and French	20.9	19.0	91	271.7	242.7	89
Neither English nor French	3.4	2.7	81	53.0	44.0	83
Aboriginal identity						
Any Aboriginal	81.7	68.0	83	86.6	71.7	83
North American Indian only	65.2	53.3	82	69.2	56.4	81
Métis only	15.7	14.0	89	15.1	13.4	89
Inuit only	0.1	0.1	76	0.4	0.3	83
Multiple or other Aboriginal	0.7	0.6	83	1.9	1.6	86
Non-Aboriginal	197.3	178.5	91	2,301.3	2,064.7	90
Immigrant status						
Immigrant	29.5	26.3	89	657.3	578.5	88
Non-immigrant	247.9	219.5	89	1,704.1	1,545.6	91
Non-permanent resident	1.6	0.8	47	26.5	12.4	47
One-year residential mobility						
Same address	241.2	215.1	89	2,044.3	1,846.6	90
Different Canadian address	31.5	26.5	84	286.8	246.3	86
Moved from outside Canada	2.0	1.5	75	26.9	19.6	73
Rural/Urban residence						
Rural and farm	120.5	103.5	86	395.1	353.7	90
Small population centre (1,000 to 29,999)	26.0	23.7	91	218.0	199.2	91
Medium population centre (30,000 to 99,999)	7.8	6.9	89	186.0	169.2	91
Large population centre (100,000 or more)	124.7	112.4	90	1,588.7	1,414.3	89

Linkage rate = (Census respondents linked to MHSIP or RPDB/Total number of census respondents)*100
Notes: Census-linked RPDB and MHSIP estimates based on census long-form respondents only. Excludes records pertaining to individuals living in institutions and records not linking to MHSIP or RPDB.
Sources: 2006 Census of Population; 2006/2007 Census-linked Registered Persons Database and Manitoba Health Services Insurance Plan.

DAD were achieved in the first iteration (72% or 1.52 million), which required an exact match on birth date, sex and postal code. Using postal code information from the T1PMF tax files, iterations 2 to 4 resulted in 80,000 more links (4%); iterations 5 to 28 added another 68,000 (3%).

Comparison of approaches

The percentages of Manitoba and Ontario census long-form respondents linking to the DAD were similar for both approaches (Table 3). Based on the *registry* approach, 7% of Manitobans had been hospitalized in an acute-care facility; based on the *non-registry* approach, the figure was 6%. In Ontario, the rate for both approaches was 5%.

As expected, the linked data reflected differential use of hospital services. For example, regardless of province or linkage approach, a higher percentage of females than males had been hospitalized. Children younger than age one and seniors were more likely than other age groups to link to hospital records. Other characteristics correlated with age and/or disability, such as not being in the labour force, were associated with higher rates of hospitalization.

Sensitivity and specificity

The majority of long-form census records had the same DAD linkage outcome, regardless of approach. The sensitivity and specificity of the Manitoba linkages were 87.9 % and 98.8%, respectively; the corresponding Ontario figures were 89.4% and 99.6% (Table 4).

The consistency of HINs among census respondents who linked to the DAD with both approaches was also high. Virtually all (99%) of the 24,487 Manitoba and 106,968 Ontario long-form census respondents who linked to the DAD according to both approaches linked to the same HIN in each approach.

Coverage evaluation

Coverage rates for all-cause hospitalizations in both provinces were comparable for the *registry* and *non-registry* linkages. Regardless of approach, the unweighted coverage rates were 23% in Manitoba and 17% in Ontario (Table 5). The weighted coverage rates were also similar: in both Manitoba and Ontario, 84% for the *registry* approach and 82% for the *non-registry* approach.

Coverage rates varied by age. Unweighted rates for Manitobans aged 75 or older were 5 or 6 percentage points below the all-ages total. For Manitoba children aged 1 to 4, weighted rates were 18 (*registry*) and 16 (*non-registry*) percentage points below the all-ages total. In both Manitoba and Ontario, and according to both approaches, weighted rates for 20- to 24-year-olds were lower (6 to 7 percentage points) than the all-ages total.

For cause-specific hospitalizations, unweighted coverage rates were closer to their Ontario and Manitoba targets than were weighted coverage weights. Unweighted cause-specific coverage rates also tended to be more similar across approaches than were those calculated with weights.

Table 2
Keys for deterministic hierarchical exact matching of 2006/2007 Discharge Abstract Database (DAD) to 2006 Census and cumulative percentage and number of DAD records linked, by iteration, Canada excluding Quebec

		DAD keys linked	
Description of key		Cumulative %	Cumulative number
Iteration	**Postal codes used complete**		
1	birth date ll sex ll census postal code FSA LDU	72.2	1,521,386
2	birth date ll sex ll postal code from 2005 T1PMF FSA LDU	74.7	1,572,846
3	birth date ll sex ll postal code from 2006 T1PMF FSA LDU	75.9	1,597,970
4	birth date ll sex ll postal code from 2007 T1PMF FSA LDU	76.1	1,601,710
	Postal codes used incomplete		
5	birth date ll sex ll census postal code FSA LDx	76.9	1,619,231
6	birth date ll sex ll census postal code FSA LxU	78.3	1,649,492
7	birth date ll sex ll census postal code FSA xDU	78.6	1,655,680
8	birth date ll sex ll census postal code FSx LDU	79.3	1,670,817
9	birth date ll sex ll census postal code FxA LDU	79.4	1,672,933
10	birth date ll sex ll census postal code xSA LDU	79.6	1,676,645
11	birth date ll sex ll postal code from 2005 T1PMF FSA LDx	79.7	1,677,620
12	birth date ll sex ll postal code from 2005 T1PMF FSA LxU	79.8	1,679,792
13	birth date ll sex ll postal code from 2005 T1PMF FSA xDU	79.8	1,680,135
14	birth date ll sex ll postal code from 2005 T1PMF FSx LDU	79.9	1,681,110
15	birth date ll sex ll postal code from 2005 T1PMF FxA LDU	79.9	1,681,220
16	birth date ll sex ll postal code from 2005 T1PMF xSA LDU	79.8	1,681,572
17	birth date ll sex ll postal code from 2006 T1PMF FSA LDx	79.8	1,682,286
18	birth date ll sex ll postal code from 2006 T1PMF FSA LxU	79.8	1,683,627
19	birth date ll sex ll postal code from 2006 T1PMF FSA xDU	80.0	1,683,846
20	birth date ll sex ll postal code from 2006 T1PMF FSx LDU	80.0	1,684,437
21	birth date ll sex ll postal code from 2006 T1PMF FxA LDU	80.0	1,684,534
22	birth date ll sex ll postal code from 2006 T1PMF xSA LDU	80.0	1,684,744
23	birth date ll sex ll postal code from 2007 T1PMF FSA LDx	80.0	1,685,177
24	birth date ll sex ll postal code from 2007 T1PMF FSA LxU	80.1	1,686,116
25	birth date ll sex ll postal code from 2007 T1PMF FSA xDU	80.1	1,686,244
26	birth date ll sex ll postal code from 2007 T1PMF FSx LDU	80.1	1,686,711
27	birth date ll sex ll postal code from 2007 T1PMF FxA LDU	80.1	1,686,767
28	birth date ll sex ll postal code from 2007 T1PMF xSA LDU	80.1	1,686,973

Note: Postal codes have six digits. The first three—FSA—pertain to the Forward Sortation Area; the last three—LDU—pertain to the Local Delivery Unit. An "x" in a postal code (FSA LDU) indicates the digit dropped from linkage key.

Discussion

The *registry* approach linked approximately 90% of long-form census respondents to the provincial health insurance registries, thereby allowing subsequent linkage to the DAD based on HINs. This rate is high, given that large-scale Canadian studies have used thresholds of about 75% as the point at

Table 3
Number and percentage of long-form 2006 Census respondents discharged from acute-care hospitals according to registry and non-registry linkage, by selected socio-demographic characteristics, Manitoba and Ontario, 2006/2007

	Census respondents discharged from hospital at least once							
	Manitoba				Ontario			
	Registry		Non-registry		Registry		Non-registry	
	'000	%	'000	%	'000	%	'000	%
Total	**18.4**	**7**	**17.9**	**6**	**118.8**	**5**	**115.7**	**5**
Sex								
Male	6.8	5	6.6	5	46.6	4	45.1	4
Female	11.6	8	11.3	8	72.1	6	70.5	6
Age group								
Younger than 1	0.6	15	0.7	16	4.0	15	3.9	15
1 to 4	0.4	3	0.4	3	2.6	2	2.4	2
5 to 9	0.3	2	0.3	2	1.7	1	1.6	1
10 to 14	0.4	2	0.4	2	1.9	1	1.8	1
15 to 19	1.2	6	1.2	5	3.5	2	3.3	2
20 to 24	1.4	7	1.4	7	5.3	3	5.1	3
25 to 34	3.0	9	2.9	8	18.7	6	18.7	6
35 to 44	2.0	5	1.9	5	13.1	3	13.2	3
45 to 54	2.0	5	1.9	5	12.6	3	12.4	3
55 to 64	2.0	7	1.9	7	15.3	6	15.0	6
65 to 74	2.0	12	1.9	11	16.8	10	16.5	10
75 or older	3.1	20	2.9	19	23.3	17	21.7	16
Economic family income quintile								
1 (lowest)	4.5	8	4.4	8	25.9	5	24.4	5
2	4.6	8	4.4	8	26.3	5	25.5	5
3	3.6	6	3.5	6	23.5	5	23.1	5
4	3.1	6	3.1	5	22.0	5	21.7	5
5 (highest)	2.6	5	2.6	5	21.0	4	20.9	4
Education (aged 18 or older)								
Secondary graduation or more	8.6	6	8.5	6	76.7	5	75.7	5
Less than secondary graduation	7.4	11	7.0	11	29.8	9	28.4	8
Employment status								
Employed	6.2	5	6.2	5	46.1	4	46.2	4
Unemployed	0.9	8	0.8	8	3.5	4	3.3	4
Not in the labour force	11.4	15	10.9	14	69.2	11	66.1	10
Knowledge of official languages								
English only	17.0	7	16.5	7	102.7	5	100.1	5
French only	0.0	7	0.0	6	0.6	6	0.6	6
Both English and French	1.1	5	1.0	5	11.9	4	11.7	4
Neither English nor French	0.3	8	0.3	8	3.5	7	3.3	6
Aboriginal identity								
Any Aboriginal	7.0	9	6.8	8	5.8	7	5.4	6
North American Indian only	6.0	9	5.8	9	4.8	7	4.5	6
Métis only	1.0	6	0.9	6	0.8	6	0.8	5
Inuit only	0.0	7	0.0	7	0.0	4	0.0	5
Multiple or other Aboriginal	0.1	7	0.0	7	0.1	6	0.1	6
Non-Aboriginal	11.4	6	11.1	6	113.0	5	110.3	5
Immigrant status								
Immigrant	1.8	6	1.7	6	34.2	5	33.1	5
Non-immigrant	16.6	7	16.1	6	84.1	5	82.1	5
Non-permanent resident	0.0	2	0.0	2	0.5	2	0.5	2
One-year residential mobility								
Same address	15.7	7	15.3	6	99.9	5	97.8	5
Different Canadian address	2.0	6	1.9	6	13.9	5	13.1	5
Moved from outside Canada	0.1	3	0.1	3	0.9	3	0.9	3
Rural/Urban residence								
Rural and farm	9.4	8	9.3	8	21.7	5	22.0	6
Small population centre (1,000 to 29,999)	2.0	8	2.0	8	13.8	6	13.5	6
Medium population centre (30,000 to 99,999)	0.5	6	0.5	6	11.1	6	10.6	6
Large population centre (100,000 or more)	6.5	5	6.1	5	72.2	5	69.6	4

Notes: % hospitalized = (number of long-form census respondents hospitalized/number of long-form census respondents) * 100.
Sources: 2006 Census of Population; Registry and non-registry census-linked 2006/2007 Discharge Abstract Database (DAD).

Table 4
Sensitivity (true positives) and specificity (true negatives) of non-registry-based in relation to registry-based linkage results, Manitoba and Ontario, 2006/2007

Manitoba				Ontario			
	Registry (gold standard)				Registry (gold standard)		
	Linked to DAD				Linked to DAD		
Non-registry	Yes	No	Total	Non-registry	Yes	No	Total
Linked to DAD	a = true positive	b = false positive		Linked to DAD	a = true positive	b = false positive	
Yes	28,428	2,848	31,276	Yes	106,968	9,366	116,334
No	c = false negative	d = true negative		No	c = false negative	d = true negative	
	3,898	243,763	247,661		12,712	2,258,865	2,271,577
Total	32,326	246,611	278,937	Total	119,680	2,268,231	2,387,911
Sensitivity a/(a+c) = 87.9% (% true positives)				Sensitivity a/(a+c) = 89.4% (% true positives)			
Specificity d/(b+d) = 98.8% (% true negatives)				Specificity d/(b+d) = 99.6% (% true negatives)			

Sources: Registry and non-registry census-linked 2006/2007 Discharge Abstract Database (DAD).

which linked data are deemed appropriate for research applications.[8,15,26] Consistent with other studies, this analysis shows that name-based linkages (*registry* approach) yield slightly higher linkage rates than do non-name-based linkages (*non-registry* approach), but that both are sufficient for research.[1]

Many characteristics associated with lower linkage rates in this analysis have been reported previously.[5,42-44] Linkage rates were relatively low for individuals with lower socio-economic status, those identifying as Aboriginal, people without knowledge of Canada's official languages, rural residents, and people who recently moved to Canada.

The results of the *registry* and *non-registry* linkages to the DAD were similar: 5% of Ontario and 6% to 7% of Manitoba census respondents linked to the hospital data. As well, linked data were consistent with expected patterns of hospital use in that higher percentages of the poor, the elderly, and people identifying as Aboriginal had been hospitalized.[45,46] This suggests that the *non-registry* approach can yield a research-quality dataset.

The coverage evaluation demonstrated consistency across linkage approaches. Unweighted coverage rates were higher in Manitoba than in Ontario, reflecting the higher percentage of Manitoba's population who completed a long-form census questionnaire and higher rates of hospitalization in Manitoba.[47] When weights were applied, the Manitoba and Ontario coverage rates often became more similar.

The linked data tended to underestimate hospitalizations of seniors, children aged 1 to 4, and people aged 15 to 44. To some extent, this is because the population represented in the long-form census data does not exactly match the groups captured by hospital data. For example, the linked data do not include the institutionalized population, who are partially included in the hospital data. Lower coverage rates of younger adults may be related to census undercoverage of populations with less stable living arrangements and/or incomplete coverage of some Aboriginal people in Ontario.[48,49]

Limitations

The linked data have several limitations. Because this study concerns only two provinces, the generalizability of the results to other jurisdictions is unclear. Preliminary analyses of coverage rates suggested potential difficulties achieving statistical significance when covariates are too narrowly defined. The results showing that coverage sometimes worsened when weights were applied indicate that the use of census weights should be considered on a study-by-study basis.

What is already known on this subject?

- Record linkage is a cost-effective way to add value to administrative and survey datasets.
- Combining complementary data sources permits analyses that would otherwise not be possible.
- The use of provincial health insurance registries to facilitate record linkages is well established.
- It is less clear if linkages based on birth date, sex and postal code can produce research-quality data.

What does this study add?

- This study compares registry and non-registry approaches to linking census data to hospitalization records in the Discharge Abstract Database.
- The two approaches produced similar results.
- The results suggest that the non-registry approach is a viable alternative in the absence of access to provincial health insurance registry data.

Table 5
Unweighted and weighted coverage rates for acute care hospital discharges, all-cause and selected most responsible diagnoses, by linkage approach, Manitoba and Ontario, April 1, 2006 through March 31, 2007

	Manitoba					Ontario				
	Number of hospital discharges (unlinked)	Unweighted		Weighted		Number of hospital discharges (unlinked)	Unweighted		Weighted	
		Registry	Non-registry	Registry	Non-registry		Registry	Non-registry	Registry	Non-registry
	'000	%		%		'000	%		%	
All causes	113.8	23	23	84	82	917.7	17	17	84	82
Sex										
Male	45.7	22	22	81	80	381.8	17	17	84	83
Female	68.1	24	24	86	83	535.8	17	17	84	83
Age group										
Younger than 1	2.9	23	26	75	77	25.7	16	16	81	78
1 to 4	2.0	28	28	65	66	20.1	18	17	86	82
5 to 9	1.4	29	29	79	79	11.6	18	17	86	84
10 to 14	1.7	29	28	80	78	12.3	18	17	88	85
15 to 19	4.7	34	32	87	82	24.7	17	16	79	77
20 to 24	7.2	27	27	77	75	37.9	16	15	77	75
25 to 34	15.9	24	24	80	79	127.4	16	16	82	82
35 to 44	10.6	26	25	86	86	94.9	17	17	83	84
45 to 54	11.7	24	24	85	83	95.6	17	17	84	85
55 to 64	13.7	22	22	83	83	118.5	18	18	87	86
65 to 74	14.7	22	22	89	87	135.9	18	17	88	86
75 or older	27.3	18	17	86	82	212.1	16	16	83	78
Circulatory system diseases (45 or older)	12.3	22	21	90	88	121.1	17	17	86	84
Sex										
Men	7.0	21	21	89	88	71.0	17	17	87	86
Women	5.3	22	22	92	88	50.1	17	16	84	81
Age group										
45 to 54	1.3	24	24	80	82	14.1	17	17	82	85
55 to 64	2.3	25	24	91	89	24.1	18	18	89	89
65 to 74	2.9	23	23	92	93	30.5	18	18	90	89
75 or older	5.8	19	18	90	86	52.4	17	16	83	79
Injury and poisoning	9.3	24	23	79	76	75.1	16	16	81	78
Sex										
Men	5.0	24	23	76	73	37.9	16	16	80	78
Women	4.4	24	23	82	80	37.2	17	16	82	78
Age group										
0 to 14	0.8	31	31	76	80	5.8	18	17	84	82
15 to 24	1.3	30	30	82	77	7.6	16	16	77	76
25 to 44	2.1	24	24	62	61	13.9	15	15	72	71
45 to 64	2.2	24	23	87	83	18.7	17	17	86	86
65 or older	2.9	19	18	84	82	29.2	16	15	82	77
Pregnancy, childbirth and puerperium (15 to 44)	16.7	28	27	88	85	138.6	17	17	83	83
Age group										
15 to 24	5.5	32	30	82	77	24.9	15	15	73	71
25 to 34	8.9	25	25	90	88	84.6	17	17	84	85
35 to 44	2.2	26	27	93	95	29.1	17	17	86	87

Coverage rate = (Total hospitalizations of of long-form census respondents/Total hospitalizations according to DAD)*100
Notes: Unlinked and linked DAD estimates estimates exclude stillbirths, non-acute care, non-Canadians, birth dates after Census day (May 16, 2006), health cards not issued by Manitoba or Ontario (only unlinked), and patients transfered from seniors' residences.
Sources: 2006 Census of Population; 2006/2007 census-linked Discharge Abstract Database (DAD).

Conclusion

The comparison of linkage approaches provides evidence that research-quality data can be produced without recourse to provincial health insurance registries, most of which are not available to Statistics Canada. The research opportunities offered by the non-registry linked file are great due to the nationally representative sample and the statistical power provided by its size and population coverage. Nonetheless, users of linked data should consider the impact of the linkage methodology, the linkage and coverage rates, and population exclusions on their analyses. ■

References

1. Winkler WE. Record linkage. *Sample Surveys: Design, Methods and Applications* 2009; 29A: 351-80.
2. Lix L, Yogendran M, Burchill M, et al. *Defining and Validating Chronic Diseases: An Administrative Data Approach*. Winnipeg: Manitoba Centre for Health Policy, 2006.
3. Iron K, Zagorski BM, Sykora K, et al. Living and dying in Ontario: An opportunity for improved health information. *ICES Investigative Report*. Toronto: Institute for Clinical Evaluative Sciences, 2008.
4. Leong VW, Morgan S, Wong ST, et al. Registration for public drug benefits across areas of differing ethnic composition in British Columbia, Canada. *BMC Health Services Research* (electronic) 2010; 10(171). Available at: http://www.biomedcentral.com/1472-6963/10/171
5. Wilkins R, Tjepkema M, Mustard C, et al. The Canadian census mortality follow-up study, 1991 through 2001. *Health Reports* 2008; 19(3): 25-43.
6. Wilkins K, Shields M, Rotermann M. Smokers' use of acute-care hospitals—A prospective study. *Health Reports* 2009; 20(4): 75-84.
7. Holman CD, Bass AJ, Rosman DL, et al. A decade of data linkage in Western Australia: Strategic design, applications and benefits of the WA data linkage system. *Australian Health Review* 2008; 32(4): 766-77.
8. Ng E. The healthy immigrant effect and mortality rates. *Health Reports* 2011; 22(4): 24-30.
9. Hwang SW, Wilkins R, Tjepkema M, et al. Mortality among residents of shelters, rooming houses, and hotels in Canada: 11-year follow-up study. *British Medical Journal* 2009; 339: b4036. doi:101136/bmj.b4036.
10. Tjepkema M, Wilkins R. Remaining life expectancy at age 25 and probability of survival to age 75, by socio-economic status and Aboriginal ancestry. *Health Reports* 2011; 22(4): 31-6.
11. Manuel DG, Perez R, Bennett C, et al. *Seven More Years: The Impact of Smoking, Alcohol, Diet, Physical Activity, and Stress on Health and Life Expectancy in Ontario*. An ICES/PHO Report. Toronto: Clinical Evaluative Sciences and Public Health Ontario, 2012.
12. Li B, Quan H, Fong A, et al. Assessing record linkage between health care and Vital Statistics databases using deterministic methods. *BMC Health Services Research* (electronic) 2006; 6(48). Available at: http://www.biomedcentral.com/1472-6963/6/48
13. Karmel R, Anderson P, Gibson D, et al. Empirical aspects of records linkage across multiple data sets using statistical linkage keys : the experience of PIAC cohort study. *BMC Health Services Research* 2010; 10:41. Available at: http://www.biomedcentral.com/1472-6963/10/41
14. NCSIMG. *Statistical Data Linkage in Community Services Data Collection*. Canberra: Australian Institute of Health and Welfare, 2004. Available at : http://www.aihw.gov.au/WorkArea/DownloadAsset.aspx?id=10737419651&libID=10737419650
15. Fransoo R, Martens P, Prior H, et al. *Adult Obesity in Manitoba: Prevalence, Associations, and Outcomes*. Winnipeg, Manitoba: Manitoba Centre for Health Policy, 2011.
16. Bernier J, Nobrega K. Overview of record linkage. *SYMPOSIUM 99: Combining Data from Different Sources, May 1999*. Ottawa: Minister of Industry, 2000.
17. Australian Institute of Health and Welfare. *Comparing an SLK-based and a Name-based Data Linkage Strategy: An Investigation into the PIAC Linkage*. Data Linkage Series No. 11 (Catalogue CSI 11) Canberra: Australian Institute of Health and Welfare, 2011.
18. Hoving JL, Del Monaco A, MacFarlance E. Methological issues in linking study participants to Australian cancer registries using different methods: lessons from a cohort study. *Australian and New Zealand Journal of Public Health* 2005; 29(4): 378-82.
19. Beauchamp A, Tonkin AM, Sundararajan V, et al. Validation of de-identified record linkage to ascertain hospital admissions in a cohort study. *BMC Medical Research Methodology* 2011; 11(42). doi:10.1186/1471-2288-11-42.
20. Australian Bureau of Statistics. *Assessing the Quality of Linking School Enrolment Records to 2011 Census Data: Deterministic Linkage Methods*. Research Paper 1351.0.55.045. Available at: http://www.ausstats.abs.gov.au/ausstats/subscriber.nsf/0/FAD1F2FC23A36CE4CA257C3F00180F64/$File/1351055045_dec%202013.pdf
21. Jebamani LS, Burchill CA, Martens PJ. Using data linkage to identify First Nations Manitobans. *Canadian Journal of Public Health* 2005; 96(1): S28-32.
22. Rotermann M. Evaluation of the coverage of linked Canadian Community Health Survey and hospital inpatient records. *Health Reports* 2009; 20(1): 45-51.
23. Da Silveira DP, Artmann E. Accuracy of probabilistic record linkage applied to health databases: Systematic review. *Revista Saúde Pública* 2009; 43(5): 875-82.
24. Migowski A, Chaves RBM, Coeli CM, et al. Accuracy of probabilistic record linkage in the assessment of high-complexity cardiology procedures. *Revista Saúde Pública* 2011; 45(2): 269-75.
25. Houle C, Berthelot JM, David P, et al. Matching census database and Manitoba health care files. *Federal Committee on Statistical Methodology Papers* 1997. Available at: http://www.fcsm.gov/working-papers/choule.pdf
26. Brennan J, Lessard M. A record linkage example: Linkage of the Citizenship and Immigration (CIC) landing file to the T1FF. Statistics Canada Methodology Seminar, April 9, 2013.
27. Statistics Canada. *Approved Record Linkages*. Available at: http://www.statcan.gc.ca/record-enregistrement/summ-somm-eng.htm
28. Statistics Canada. *Directive on Record Linkage*. Available at: http://icn-rci.statcan.ca/31/31b/31b_10c025-eng.htm
29. Statistics Canada. 2006 *Census of Population*. Available at: http://www12.statcan.gc.ca/census-recensement/2006/ index-eng.cfm
30. St. Pierre M, Gagné W. Création d'un prototype de la base de la population canadienne. Statistics Canada Internal Working Documentation, 2012.
31. Ministry of Health and Long-term Care. *Resource Manual for Physicians*. Available at: http://www.health.gov.on.ca/english/providers/pub/ohip/physmanual/download/section_5.pdf
32. Roos LL, Mustard CA, Nicol JP, et al. Registries and administrative data: Organization and accuracy. *Medical Care* 1993; 31(3): 201-12.

33. Roos LL, Nicol JP. A research registry: Uses, development, and accuracy. *Journal of Clinical Epidemiology* 1999; 52(1): 39-47.

34. Canadian Institute for Health Information. *Discharge Abstract Database (DAD) Metadata*. Available at: http://secure.cihi.ca/cihi-ext-portal/internet/en/document/types+of+care/hospital+care/acute+care/dad_metadata

35. Canadian Institute for Health Information. *Data Quality of the 2006-2007 Discharge Abstract Database*. Ottawa, Ontario: Canadian Insititute for Health Information, 2009. Available at: http://secure.cihi.ca/cihiweb/products/DAD_2006-07_Main_Report%20_en_web.pdf

36. Reicker A. *Longitudinal Health Administrative Data (LHAD) Project*. Internal linkage of the Manitoba and Ontario health population files. Unpublished internal documentation, 2012.

37. Fellegi IP, Sunter A. A theory of record linkage. *Journal of the American Statistical Association* 1969; 64: 1183-210.

38. Ntwari A, St-Jean H, Hortop E, et al. *Longitudinal Health Administrative Data (LHAD) Project*. External linkages between the Ontario/Manitoba health insurance registries and Ontario/Manitoba subsets of the 2006 Census of Population. Unpublished Internal documentation, 2012.

39. Statistics Canada. *2006 Census Technical Report on Sampling and Weighting*. Ottawa: Minister of Industry, 2009. Available at: http://www12.statcan.gc.ca/census-recensement/2006/ref/rp-guides/rp/sw-ep/pdf/92-568-X2006001-eng.pdf

40. Canadian Institute for Health Information. *Health Care in Canada*. Ottawa: Canadian Institute for Health Information, 2011. Available at: http://secure.cihi.ca/freeproducts/HCIC_2011_seniors_report_en.pdf

41. Trottier H, Martel L, Houle C, et al. Living at home or in an institution: What makes the difference for seniors. *Health Reports* 2000; 11(4): 49-61.

42. Australian Institute of Health and Welfare, Australian Bureau of Statistics. *National Best Practice Guidelines for Data Linkage Activities Relating to Aboriginal and Torres Strait Islander People*. Canberra: Australian Institute of Health and Welfare, 2012. Available at: http://www.aihw.gov.au/publication-detail/?id=60129542148

43. Bass J, Garfield C. Statistical linkage keys: How effective are they? Proceedings of the Symposium on Health Data Linkage, 2003. Available at: http://www.publichealth.gov.au/pdf/reports_papers/symposium_procdngs_2003/bass.pdf

44. Blakely T, Woodward A, Salmond C. Anonymous linkage of New Zealand mortality and Census data. *Australian and New Zealand Journal of Public Health* 2000; 2491: 92-5.

45. Sanmartin C, Khan S, LHAD Research Team. Hospitalizations for Ambulatory Care Sensitive Conditions (ACSC): The factors that matter. *Health Analysis Division Working Paper Series* 2011; 7. Available at: http://www.statcan.gc.ca/pub/82-622-x/82-622-x2011007-eng.pdf

46. Carrière G, Garner R, Sanmartin C, et al. Acute-care hospitalizations and Aboriginal identity in Canada, 2001/2002. *Health Analysis Division Working Paper Series* 2010; 5. Available at: http://www.statcan.gc.ca/pub/82-622-x/82-622-x2010005-eng.pdf

47. Canadian Institute for Health Information. *Highlights of 2008-2009 Inpatient Hospitalizations and Emergency Department Visits*. Ottawa: Canadian Institute for Health Information, 2010. Available at: http://www.cihi.ca/CIHI-ext-portal/pdf/internet/NACRS_HIGHLIGHTS_18May10_EN

48. Statistics Canada. *Incompletely Enumerated Indian Reserves and Indian Settlements*. Available at: http://www12.statcan.gc.ca/census-recensement/2006/ref/notes/aboriginal-autochtones-eng.cfm

49. Statistics Canada. *2006 Census Technical Report: Coverage*. Ottawa: Minister of Industry, 2010. Available at: http://www12.statcan.gc.ca/census-recensement/2006/ref/rp-guides/rp/coverage-couverture/pdf/92-567-X2006001-eng.pdf